DOCUMENT RESUME

ED 100 744 SO 008 024

AUTHOR Scriven, Michael
TITLE Evaluating Social Studies and Citizenship Education.
INSTITUTION National Council for the Social Studies, Washington, D.C.
SPONS AGENCY Education Commission of the States, Denver, Colo. National Assessment of Educational Progress.
NOTE 20p.; A paper commissioned for Task 4; related documents are SO 008 019-026

EDRS PRICE MF-$0.75 HC-$1.50 PLUS POSTAGE
DESCRIPTORS *Citizenship; *Educational Assessment; Educational Objectives; Ethics; *Evaluation; Evaluation Criteria; Measurement Goals; *Measurement Instruments; Models; National Surveys; *Social Studies; Values
IDENTIFIERS *National Assessment

ABSTRACT

This paper was commissioned to develop new perspectives on the evaluation of the National Assessment of Education Progress (NAEP) for Citizenship and Social Studies. It criticizes the assessment model and advocates three new approaches to evaluating social studies education: a comparative approach with direct international comparisons of programs and results; direct comparison of pupil performance in schools with radically different approaches to social studies; and a sociological approach where values of various age-groups and adults are studied through their communications and decision processes. Criticism of the current NAEP model includes the restriction of goals and test items to those socially acceptable by the states, educators, and lay citizens. Schools and states should not be able to vote on the standards on which they are to be judged. The objectives and test items showed a lack of emphasis on ethics and their relation to the students' life situations. The interpretation of results was overoptomistic in view of the low scores achieved by the sample population. (DE)

ED 100744

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

EVALUATING SOCIAL STUDIES AND CITIZENSHIP EDUCATION

Michael Scriven
University of California, Berkeley

0. Introduction

There is tremendous overlap between these areas--well-illustrated by the overlap in the goals and objectives developed by the two different institutions (ETS and AIR) that contracted with NAEP to develop specific statements. I shall not make any great effort to separate them here because our concern is with evaluation models and only secondarily with specific content--and very similar problems arise in both areas.

Let's call the NAEP approach, well described by Bob Taylor, the first approach to evaluating social studies education. I shall describe three other approaches very briefly and suggest a synthesis. Then I shall look in somewhat more detail at certain features of the NAEP approach and Bob Taylor's comments. I'll begin by focusing on the area of citizenship values, understanding of law and due process, etc., because it's more important than geography and in worse shape.

SP 008 024

1. Alternatives to NAEP

A second possibility would be a comparative approach in which direct international comparisons--for example--were sought. This would require a substantial but not complete revision of the item pool in order to make really direct comparisons: but

some direct and many indirect ones can be made using the data we do have from the Institute of International Education studies and others. These comparisons are extremely important--even more than they were in the math studies, for example--in showing what can be done as well as what is being done. For the citizenship/social studies (c/ss) areas are not so abstruse that one can reasonably suppose them (or most of them) to be beyond the grasp of a substantial majority of pupils. Of course, there are important differences between countries that would make it necessary to proceed with caution, in inferring from what has been done elsewhere to what could be done here, but the differences are not such as to make attempts (experiments) absurd; which is all we could justify to start with, anyway. Of course, too, tremendous chauvinism is associated with the c/ss area, and there would be many who would condemn any attempt to match the performance of other countries per se. However, the merits of that argument seem as slight here as in, e.g., the automobile or the psychiatric or the adult education field. With GM switching to the Wankel engine, acknowledgedly copying Mercedes in styling and suspension and Ferrari in design (with the 1975 models), one can hardly argue that the experts can't see any transferability of foreign ideas. Similar cases are well-known in ther other fields mentioned. If we are interested in reducing the level of anti-social activities and basic ignorance about constitutional and other rights, and about human nature--which is what c/ss is all about--then it seems appropriate to look for possible improvements wherever we can find

them. There is plenty of evidence in the comparative educational studies to date to suggest that we could do better; but we do need more precise comparisons. Telling us whether we're doing as well as we could is one of the functions of evaluation, and it isn't done very well by NAEP.

Another type of comparison that would really be significant would involve comparisons between the performance of pupils--still measured mostly on paper and pencil tests, as in NAEP--attending schools with radically different approaches to c/ss. The contrast between those with a conventional curriculum and those using some of the alternative approaches, e.g., Project Social Studies materials, might be illuminating; and the discovery that there wasn't a contrast would also be illuminating. Of course, NAEP does make some comparisons, e.g., between performance of black and white pupils. But, it's a little hard to alter people's color; it's less hard to introduce new curricula or methods. One might put it this way: NAEP, and most state assessment programs, are pretty good photographers, but not very good buying guides. For that you need the relevant comparisons, viz., those between the available options.

A third possibility involves switching to a very different kind of item, albeit still a paper and pencil (or vocal) test. Here we'd go to something like the Social Issues Analysis Test, where the item might present a page-long newspaper editorial or a dialogue and a series of rather searching questions on it, of a wide range of difficulty. We might allow a lot of time

for this--perhaps an hour. An essential selection criterion for items would be novelty: typically they would be hypothetical cases to which the "approved" answer has not yet been identified by adults with whom the pupils interact (to rule out parrotting). The reasons for shifting to this kind of item are (a) that the few items like this that have been released reveal the most appalling incompetence in operating with the simplest constitutional or moral principles (e.g., freedom of speech) and it is now most urgent to clarify the real situation; (b) although the analysis of responses would require formidable training and talent on the part of the scorers (since much reading between the lines would be necessary) there would be a corresponding increase in the significance of the results. Instead of telling us where the pupils are at, this kind of test can tell us where they are capable of moving to, which is our only hope (given the abysmal level of performance at the moment). For the page of dialogue can involve argumentation, and can call for evidence of understanding the steps in the argument and their effect on the reader (e.g., by using interspersed questions and indelible markers, etc.). Despite the use of hypothetical situations, the responses are much more likely to be realistic here than in the present items, where the use of stereotypes by the student can easily provide a facade of answers that tell us nothing about the probable response to a new case. In short, bad though the present answers are, they may well give huge overestimates of the merit of the respondents, which leads us back to reason (a) above.

A somewhat radical extension of this approach leads to a fourth evaluation, which would move into the field and away from pencil and paper tests, and use the best skills of the anthropologist and the sociologist (besides those of an extremely acute content analyst) to identify the values of various age-groups and adults in our society from a study of their communications and decision-processes. To take an extreme--but extremely important--example; during the hearings before the Ervin committee we were presented with a very detailed picture of the level of moral analysis and citizenship behavior of the White House staff. The addition of the tapes has made this a very complete data source for the kind of question I'm raising here. Similar analyses can be done of the discussion at the school board and in the local press of a proposed decriminalization order to a city police department (and of subsequent events), or of the discussions in an eighth grade classroom of a proposal to vest disciplinary powers in the students (and of the subsequent events). They are tricky; there are few analysts presently equipped to deal with them objectively--but oh, what a treasure trove for the evaluator is there! For here we can bypass the problem of test invalidity; here we are dealing with real actions and perceptions. Despite the massive media coverage of Watergate, I never saw any analysis of the significance of the conceptions revealed by Haldeman and Ehrlichman on the stand. Most people got the feeling that they were "sort of morally blind", that they were abusing their power. But consider Ehrlichman's justification

of the burglary of Dr. Fielding's office; "as we saw it, it was as if you learned that a map was stored in the vault of a D.C. bank, that showed the location of a bomb that would blow up the whole of the district the next day; wouldn't you think it was justifiable to 'break and enter'?" (paraphrase). (There were a dozen similar examples.) I think there is more information about the evaluation of c/ss in U.S. schools and homes in those passages than in all the corresponding test results from NAEP. This was not just one aberrant lawyer speaking. This was a line of argument--grotesquely irrelevant though it was--that readily persuaded almost everyone on whom it was tried by someone coming from the White House. We did not learn about Watergate from someone who had a better education in c/ss than Ehrlichman; we learned about it from a black nightwatchman doing his job well--for which he was essentially blacklisted.

There is a recurrent tone in the Watergate discussions at every level--media, Congress, and neighborhood--and the same note can be detected in the discussions of any other widely-discussed moral issue of our time, such as drug law and enforcement, "excess" profits by oil companies, etc. That tone is naiveté--and from our point of view, particularly naiveté about the psychological nature of mankind, society, and morality. We need more careful evaluation than we have yet had to determine whether this impressionistic reaction is ill-based or not. The third and fourth methods described here use simple enough procedures, which we have often applied in evaluating competency in other areas of interest, e.g., in testing cognitive, mechanical, and

administrative skills. I believe they deserve more serious application in c/ss, where we have so far--with regard to our own society--alternated between oversimplified paper and pencil tests and over-emotional social documentary.

One feature of the field-study or anthropological approach which deserves some stress is that it does not begin (or does not need to begin) with the massive effort involved in developing goals. There is something slightly inappropriate about that effort for an evaluation task, it seems to me; it is exactly the right activity for developing a new curriculum, but that is hardly what NAEP was supposed to be up to. (It's perhaps not too surprising that considerable opposition to NAEP arose from those who felt that it was attempting to impose a monolithic c/ss curriculum on U.S. schools; the complaint might seem stupid at first glance, but on second thought reflects some sensitivity to a significantly possible outcome, school politics being what they are.) It seems plausible enough to argue that you can't set up tests until you know what they are tests for, and what they are tests for, i.e. the goals of c/ss education. But that's an error, as we'll see in the next section. Here I'll just stress the existence of an alternative approach. One could have had a team analyzing adult behavior in the citizenship area for deficiencies by identifying the optimal feasible behavior in the situation in question and extracting the discrepancies. After a long search, one would then classify the discrepancies and set up the assessment program to determine the extent of these

deficiencies in the population. This involves no reference to the goals of c/ss education, though such could be inferred from it; it short-circuits that concept.

So the three models I am proposing might be called the comparative, the simulation, and the anthropological models. They are mere sketches here, of course, but I believe they do serve to open our minds to the existence of rather different approaches to evaluation of c/ss education--possibly they will serve as useful targets for discussion. We have become somewhat fixated on the "standard" model of assessment, and we have invested in it very heavily (see annual reports of ETS Center for Statewide Assessment). I think we have become too rigid in using this model, and I see no reason why some diversion of resources could not be made to include at least some of the other models I have described.

But that's not the only possible way to change. There are major changes in the NAEP model that deserve consideration, and that could also produce an "alternate form" which could be used alongside the (desirable for obvious reasons) continued use of the present forms. We'll turn to these in the next section.

2. Changes in NAEP

A tremendous price was paid for "political" acceptability of the NAEP approach, and this may well have been the right decision. However, there is some point in talking about ideal ways of evaluating, and even the feasibility question is probably due for reconsideration. The two big trade-offs (or sell-outs, depending on how radical one feels this morning) are:

(a) restriction of goals to those "accepted as an educational task by the school"[1]; or "acceptable to most educators and considered desirable teaching goals in most schools."[2] A further restriction was to goals that were "considered desirable by thoughtful lay citizens."[3]

(b) restriction of items to those which most states liked. In some cases, it is clear, particular states would not accept certain items in the version of the test forms used within their boundaries, and to avoid this becoming widespread, compromises had to be made on other items. Of course, c/ss were the areas hit hardest by this constraint.

These are serious limitations indeed. If schools and states can vote on the standards by which they are to be judged, we are simply going to lose some very fundamental criticisms. The process actually gave the veto power to each of three groups--scholars, educators, and laypeople. That's a pretty tough obstacle race for an objective to get through and some pretty crucial ones didn't make it, especially--unfortunately--those which would most acutely test the moral sensitivity of students on controversial issues.

These restrictions might be relaxed after new consultations; or they might be bypassed using the field study approach described earlier. One way or another, they are barriers to a full evaluation.

One does not judge the education of lawyers or doctors by asking the law school or medical school for criteria (or letting them veto external lists). One judges it by a careful analysis

of the performance of the professionals in the field, using the testimony of clients and co-workers who see that work; an analysis that looks not only at deficiencies, which will always be with us, but also at the question whether these deficiencies are the kind that could have been removed by education, preferably an education which is fiscally and temporally realistic.

The schools are permeated by a number of unfortunate ideologies in the c/ss area, ideologies which are tremendously destructive to reasonable c/ss education, and which are completely fallacious. In the light of these ideologies, educators of course reject certain kinds of goals for c/ss; yet these include many goals of the greatest importance for c/ss education. A couple of examples may suffice to illustrate the point. The fact-value distinction and the associated ideology of value-free science, is pervasive among educators (and many scholars and thoughtful laypeople). Hence they will not accept goals which assert the objectivity and factuality of certain moral standards, and the falsity of others. Indeed they go further and require (i.e., did accept) objectives like SSIIA, 17-A: "Distinguish among definitional, value, and factual issues in a dispute."[4] This is of course the thin edge of the relativistic wedge. If one can't say that it's a <u>fact</u> that Ehrlichman <u>improperly</u> approached Judge Byrne, then ethics is indeed a travesty; but of course it <u>is</u> a value claim and if these are exclusive categories, it can't be both at once.

There are other glaring omissions in the objectives lists concerning the foundations of ethics, the relation of ethics to

religion, to conscience, to the law, to custom and convention, to pragmatic considerations--the very issues on which a person's ethical commitment founders in the tempest of a personal crisis. But as my second example, let me take something less philosophical, more specific--the understanding of Communism. Is there a more important issue? Is there a worse-taught issue? Is there an issue on which we need information more desperately? Are there searching questions aimed at discovering true understanding rather than slogan-memorizing? Clearly not. Here is a case where the label on the package will pass the educators ("teaching about Communism" is an acceptable goal), but the only sane way to do it (use Communist documents and speakers, [live, taped, or filmed] as well as critical commentaries) is entirely unacceptable. The same applies to homosexuality, adultery, prostitution, violence, abortion, pornography, etc., etc.: in short, to most of the topics that are likely to produce a personal moral crisis for the graduate of or pupil in our schools, and that can be thoroughly and helpfully discussed there. Instead, they have to be discussed by the walking wounded in later life, too late for primary prophylaxis.

The second major weakness in the NAEP approach lies in the conceptualization of the goals and objectives. Without detracting from the very considerable merits of ETS and AIR, who did the work, they leave a great deal to be desired and bear the heavy signs of committee authorship. A few examples from the Citizenship goals will indicate the kind of problem that exists.

Goal A is "Show concern for the welfare and dignity of others".[5] Of course, showing concern is not what we want; we want having concern. It's attractive to go for the "behavioral objective" formulation, but it focuses on external signs when we want something much deeper. Someone who does not show concern but who gets the ambulance is better than someone who weeps hysterically.

Objective G-1 is "Try to inform themselves on socially important matters and to understand alternative viewpoints".[6] Is the goal trying? Or is the goal succeeding? Suppose you find that everyone in the U.S. K-12 system is trying to inform themselves about something, but--e.g., because of incompetent teachers--failing dismally. Would you feel that c/ss education was succeeding? This is not a semantic issue. I suspect that, in some feeble sense, most people "try to understand" the use of bloody and destructive violence by political revolutionaries in this country. I think most of them (would say they) fail. I think that shows something about the gross inadequacies of c/ss education, not something about its success. They know nothing of the philosophy of violence; they could identify none or at most one of the half-dozen powerful reasons for the use of violence; it does not even occur to them that their own country was founded on violence and has perpetrated and institutionalized violence to a massive extent. They are examples of the failure of c/ss education, and an evaluation should so identify them.

The next major failure of the NAEP effort lies in interpretation, and it really falls under two sub-headings:

interpretations by staff and interpretations by consultants whose report was published by NAEP.

Here's an example of absurdly poor staff interpretation: "One indication that students do weigh alternatives rationally was seen in the group participation exercises; 67%-79% at all three school ages gave a reason for a particular point of view at least once during the one half hour task".[7]

Giving a reason may be aimed at persuading _others_ or _rationalizing_ one's own decision; hence it is simply improper to take it as an indication of rational deciding. That error shows a very serious lack of understanding of what rationality _is_, and that lack of understanding shows up frequently.

What _would_ be evidence of critical ability and of rational decision-making? A case where prior prejudice won't give the right answer, where the answer must come by inference from the given facts of the case, in short a _new_ problem case. None occur under Goal G "Approach Civic Decisions Rationally."

There's a pervasive overoptimistic bias in the interpretations. Why should one be inclined to think that young Americans' critical ability is anything less than ludicrous when a majority of nine year olds and a quarter of thirteen year olds think that a newspaper _can't be wrong_?[8] That's after six or seven years of schooling.

Interpreting the _global_ significance of the results was left to an advisory panel. I will indicate _my_ interpretation of one small part and you'll see why I think the truly horrifying implications have not generally been recognized. Even with the

data at hand, despite the many deficiencies already indicated, much more can be inferred than either staff or advisers have recognized. The conclusions are not both precise and highly probable. But policy decisions, contrary to the usual position, do not require these conditions. We operate off probabilities and possibilities, when the risk of not doing so is high; and in this area, that's surely the situation.

Let's take the respect for freedom of speech.[9] It's often mentioned that 75% or more of the thirteen year olds thought that no one on radio or TV should be allowed to say either that "Russia is better than the U.S." or that "Some races of people are better than others." or that "It is not necessary to believe in God."

What isn't so often said (though NAEP staff noted it) is that 94% drew the line at one or more of these statements as a permissible media utterance (i.e., only 6% thought all were utterable). And the seventeen year olds still show almost 80% refusing to allow all three. And the young adults still show 68% standing four-square against freedom of speech in these medium-controversy examples. When asked why they thought these statements should be allowed, only 2/3 of the most stalwart (adult) sample could think of freedom of speech or ideas, etc., as a justification, so one should perhaps quote as the most significant statistic, the 76% of the adults who failed this simple test, treated as a simple recognition test of a well-known principle. Now how many of the reamining 24%, if on the board

of a broadcasting station, would actually stick to their verbal endorsement of this principle? The evidence (from Hartshorne and May on) suggests that it will be far fewer--perhaps only 10% instead of 24%, perhaps only 2%. And are these examples extreme tests? On the contrary. Suppose the third quote was, not "It is not necessary to believe in God", but "Belief in God is a sign of weak-mindedness, and the source of most war and cruelty", would we really have 20% left to count on?

Remembering that huge gap between professed moral principles and actual practice, how should we feel about a test of professed tolerance of other races under very mild stress as in A4--"being willing to have someone from another race be your dentist or doctor, live next door to you, represent you in an elected office, sit at the next table in a crowded restaurant, stay in the same motel or hotel", when we find that 43% of all age groups draw the line at one or more of these possibilities. When it comes to the day when the respondent's daughter actually wants to date interracially, one can have little confidence that half that 43% will remain with us (and I'd have to say that 10% would be a surprise).

Is it not disastrous that less than a quarter of young adults (22%) could give even one reason for and one reason against educational deferments for the draft?

Now I would also say that most of the remaining questions are routine questions about routine behavior and knowledge, and the subjects performed routinely on them. One can draw little joy or sorrow from those other responses. But on the issues

that test the capacity for crisis-handling in the citizenship domain--although the tests are weak and the inferences from test performances to real performance very shaky, the results I have quoted represent most of the questions asked (since there were very few) and surely they represent significant features of the answers.

What did the Panel of Reviewers think of these results?[10] (Remember that their reactions represent the only evaluative global synthesis effort by NAEP).

By and large, they thought the results were pretty encouraging. A black panelist (Tobe Johnson) rightly complained about the WASP standards built into some questions. Larry Metcalf saw the same point, and some other biases and cautioned us not to blame or credit the schools for the results.

But no one expressed horror at the plain ignorance and prejudice revealed here, and several expressed gratification. Evaluation results sometimes call for horror, and these ones do. As to blaming the schools, why not? There's no reason to think the schools couldn't change these results around if they tried and there's every reason to think they should try. No doubt families, communities and media are also to blame, and would also resist the effort to change. That doesn't show it can't be done, and if it can be and should be and isn't, then those who don't do it must share the blame. Communities can be changed by their schools; schools aren't petrified by communities in law though they may be in fact.

So I'd sum up my reactions to the NAEP effort as involving grave weaknesses of design and of interpretation, as well as great technical virtuosity in many dimensions.

3. Comments on the Taylor Report

Much of Taylor's excellent review is unexceptionable. I will just mention one disagreement.

Taylor says: "the assessment movement is counter to the humanizing movement in American education. It is promoting a closed rather than an open approach to curriculum." I think this is a very serious misconception. To expect schools to provide certain core learnings is not to inhibit their room for all sorts of innovation. To expect students to test well on understanding democracy is hardly inhibiting humanization!

4. Conclusion

I have tried to develop new perspectives on the evaluation of c/ss, partly by describing new models and partly by criticizing the present ones. I hope this will lead us towards more useful evaluation and more effective education in this area. Nothing in our national priorities is more important.

Footnotes

[1]National Assessment of Educational Progress, Citizenship Objectives, (Denver: Education Commission of the States, 1969), p. 3.

[2]National Assessment of Educational Progress, Social Studies Objectives, (Denver: Education Commission of the States, 1970), p. 2.

[3]Ibid.

[4]Ibid., p. 10.

[5]Citizenship Objectives, op. cit., p. 9.

[6]Ibid., p. 33.

[7]National Assessment of Educational Progress, Citizenship Report 2, (Denver: Education Commission of the States, 1970), p. 93.

[8]Ibid., p. 103.

[9]Ibid., p. 34-35.

[10]National Assessment of Educational Progress, Report 2 - Citizenship: Observations and Commentary of a Panel of Reviewers, (Denver, Education Commission of the States, 1970).